1 Multivariariate Partial Derivative

Let f be a function of x_1 and x_2. Pick a fixed vector $z = [z_1, z_2]^T$. For this two-variable function, the first-order Taylor expansion (around the point z) is

$$\begin{aligned} f(x) &\approx f(z) + \left.\frac{\partial f}{\partial x_1}\right|_{x=z}(x_1 - z_1) + \left.\frac{\partial f}{\partial x_2}\right|_{x=z}(x_2 - z_2) \quad \forall x \in \mathbf{R}^2 \text{ close to } z \\ &= f(z) + \left[\begin{array}{cc} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{array}\right]_{x=z} \left[\begin{array}{c} x_1 - z_1 \\ x_2 - z_2 \end{array}\right] \\ &= f(z) + \nabla f(x)|_{x=z}^T (x - z) \end{aligned} \tag{1}$$

The term

$$\nabla f(x) = \left[\begin{array}{c} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{array}\right]$$

is called the gradient of $f(x)$, a generalization of $df(x)/dx$ in the single-variable calculus. It is a 2 by 1 column vector if $f(x)$ is a mapping from $\mathbf{R^2}$ and $\mathbf{R}$. For instance, if $f(x_1, x_2) = x_1 + 2x_2$, then

$$\nabla f(x) = \left[\begin{array}{c} 1 \\ 2 \end{array}\right]$$

Note: by convention, the gradient of a $\mathbf{R^n} \to \mathbf{R}$ mapping $f(x_1, x_2, \ldots, x_n)$ is defined as a column vector:

$$\nabla f(x) = \left[\begin{array}{c} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{array}\right]$$

as it defines a vector/direction in a vector space. A corresponding definition is the derivative

$$Df(x) = \left[\begin{array}{cccc} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \ldots & \frac{\partial f}{\partial n} \end{array}\right]$$

which is a row vector and

$$\nabla f(x) = [Df(x)]^T$$

We can generalize the above result. For instance, if $f_1(x, u) = f_1(x_1, x_2, u_1, u_2)$ then the Taylor approximation around the point $(\bar{x}, \bar{u})$ is

$$\begin{aligned} f_1(x,u) \approx f_1(\bar{x},\bar{u}) &+ \left.\left[\frac{\partial f_1(x,u)}{\partial x_1}, \frac{\partial f_1(x,u)}{\partial x_2}\right]\right|_{\substack{x=\bar{x}\\u=\bar{u}}} \left[\begin{array}{c} x_1 - \bar{x}_1 \\ x_2 - \bar{x}_2 \end{array}\right] \\ &+ \left.\left[\frac{\partial f_1(x,u)}{\partial u_1}, \frac{\partial f_1(x,u)}{\partial u_2}\right]\right|_{\substack{x=\bar{x}\\u=\bar{u}}} \left[\begin{array}{c} u_1 - \bar{u}_1 \\ u_2 - \bar{u}_2 \end{array}\right] \\ &= f_1(\bar{x},\bar{u}) + \left.\nabla_x^T f_1(x,u)\right|_{\substack{x=\bar{x}\\u=\bar{u}}} (x - \bar{x}) + \left.\nabla_u^T f_1(x,u)\right|_{\substack{x=\bar{x}\\u=\bar{u}}} (u - \bar{u}) \end{aligned}$$

Table 1: Gradient examples

$f(x)$	$\nabla f(x)$	$D\nabla f(x)$
$\|x\|_2$	$\frac{x}{\|x\|_2}$	
$\|x\|_2^2$	$2x$	
$\|x\|_2^3$	$3\|x\|_2 x$	$3\|x\|_2 + 3\frac{x^T x}{\|x\|_2}$
$y^T x$	y	
$x^T y$	y	
$\log\|x\|$	$\frac{x}{(ln 10)\|x\|^2}$	
$\frac{1}{2}\|Ax-b\|_2^2$	$A^T(Ax-b)$	

If we have another similar function

$$f_2(x,u) \approx f_2(\bar{x},\bar{u}) + \left[\frac{\partial f_2(x,u)}{\partial x_1}, \frac{\partial f_2(x,u)}{\partial x_2}\right]\Bigg|_{\substack{x=\bar{x}\\u=\bar{u}}} \begin{bmatrix} x_1-\bar{x}_1 \\ x_2-\bar{x}_2 \end{bmatrix}$$
$$+ \left[\frac{\partial f_2(x,u)}{\partial u_1}, \frac{\partial f_2(x,u)}{\partial u_2}\right]\Bigg|_{\substack{x=\bar{x}\\u=\bar{u}}} \begin{bmatrix} u_1-\bar{u}_1 \\ u_2-\bar{u}_2 \end{bmatrix}$$
$$= f_2(\bar{x},\bar{u}) + \nabla_x^T f_2(x,u)\Big|_{\substack{x=\bar{x}\\u=\bar{u}}} (x-\bar{x}) + \nabla_u^T f_2(x,u)\Big|_{\substack{x=\bar{x}\\u=\bar{u}}} (u-\bar{u})$$

Then for the $\mathbf{R}^2 \times \mathbf{R}^2 \to \mathbf{R}^2$ function

$$f(x,u) = \begin{bmatrix} f_1(x,u) \\ f_2(x,u) \end{bmatrix}$$

we have

$$f(x,u) \approx \begin{bmatrix} f_1(\bar{x},\bar{u}) \\ f_2(\bar{x},\bar{u}) \end{bmatrix} + \underbrace{\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{bmatrix}}_{\nabla_x^T f(x,u)}\Bigg|_{\substack{x=\bar{x}\\u=\bar{u}}} \begin{bmatrix} x_1-\bar{x}_1 \\ x_2-\bar{x}_2 \end{bmatrix} + \underbrace{\begin{bmatrix} \frac{\partial f_1}{\partial u_1} & \frac{\partial f_1}{\partial u_2} \\ \frac{\partial f_2}{\partial u_1} & \frac{\partial f_2}{\partial u_2} \end{bmatrix}}_{\nabla_u^T f(x,u)}\Bigg|_{\substack{x=\bar{x}\\u=\bar{u}}} \begin{bmatrix} u_1-\bar{u}_1 \\ u_2-\bar{u}_2 \end{bmatrix}$$
$$\triangleq f(\bar{x},\bar{u}) + A(x-\bar{x}) + B(u-\bar{u})$$

From here we learnt how to compute the derivative and gradient of a multi-input multi-output function:

$$D_x \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{bmatrix}, \ \nabla_x \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} \end{bmatrix}$$

Similar to single-variable calculus, we can use the chain rule to compute multi-variable derivatives:

$$Dg(f(x)) = D_g g(f(x))\, Df(x)$$

Hence

$$\nabla g(f(x)) = \nabla f(x)\, \nabla_g g(f(x))$$

The gradient of some common functions are listed in the following table.

Exercise: Prove the results in Table 1.

Assume compatible dimensions. Show that

$$\nabla_X Tr\,(XY) = Y^T$$

where $Tr(XY)$ denotes the trace of the matrix XY. Hint:

$$\frac{\partial Tr\,(XY)}{\partial X_{ij}} = \mathrm{Y}_{ji}$$